



Title	内容分析から計量テキスト分析へ -継承と発展をめざして-
Author(s)	樋口, 耕一
Citation	大阪大学大学院人間科学研究科紀要. 2006, 32, p. 1-27
Version Type	VoR
URL	<a href="https://doi.org/10.18910/11920">https://doi.org/10.18910/11920</a>
rights	
Note	

*The University of Osaka Institutional Knowledge Archive : OUKA*

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

# 内容分析から計量テキスト分析へ

—継承と発展を目指して—

樋 口 耕一

## 目 次

1. はじめに
2. 内容分析
  - 2.1 概略と定義
  - 2.2 質的データの計量的分析
  - 2.3 コンピュータ利用のありかた
3. 国内での展開
  - 3.1 国内における内容分析とコンピュータ利用
  - 3.2 計量テキスト分析
4. 計量テキスト分析の拡充
  - 4.1 定義
  - 4.2 具体的な方法



## 内容分析から計量テキスト分析へ －継承と発展を目指して－

樋口 耕一

### 1. はじめに

内容分析 (content analysis) とは、社会調査によって得られる質的データを分析する方法である。この方法は我が国でも早くから紹介されているものの（池内 1954, 1956）、その後の普及・発展が十分に進んでいるとは言い難い。内容分析では多くの場合、データをいくつかのカテゴリーに分類した上で、各カテゴリーに分類されたデータの数を数え上げたりといった、計量的分析が行われる。その際、「分類基準作成が非常に困難で、その上、手数がかかるわりに報いられる点が少なく、一時の思いつきでは意味がなく、長期間の送り手の傾向を見るのが目的なので、だれでもが簡単にできるものではない」（加納 1961: 16）という印象を持たれたようだ。また、我が国では欧米に比して、内容分析を用いた研究そのものが少ないという報告もある（三上 1988）。

その一方で、どちらかというと商業的・実務的な局面での利用が多いようだが、「テキストマイニング」と呼ばれる方法が現在注目を集めしており、次々と概説書が公刊されている（石井 2002; 山崎・松田 2002; 藤井ほか 2005）。これは、質的データの中でも特にテキスト型（文章型）データを分析する方法で、情報科学の分野で提案されたものである。テキストマイニングでは、コンピュータによってデータ中から自動的に言葉が取り出され、様々な統計手法を用いた探索的な分析が行われる。それによってパターンやルール、ひいては知識の発見を目指すのだという。これを実際に用いるためのテキストマイニング用ソフトウェアについても、1,000万円を超えるものから無料のものまで、様々なものが国内でも販売・頒布されている。国内では、内容分析の概説書が90年代以降1冊も刊行されておらず<sup>1)</sup>、内容分析用と銘打って販売されているソフトウェアも見あたらないという状況に比して、テキストマイニングの状況はまさに対照的である。また、テキストマイニングの方法が語られる際に、内容分析がほとんど取り上げられないということも、内容分析の普及の程度を表しているよう。

しかし、普及が進んでいないからといって、内容分析は役目を終えた過去の遺物では

ない。もちろん、テキストマイニングにおいて活用されているような、コンピュータやそれを用いた自然言語処理・統計解析の進歩は無視できない。だが、こうした技術の進歩が、テキスト型データの分析におけるあらゆる問題を解決してくれるのだろうか。自動要約が目標であれば話は別だが、分析という営みには人間の判断が必然的に含まれる。当然そこには、客観性ないし信頼性の維持という問題や、質的なデータを量的に分析することで得られるものと失うものとの兼ね合いという問題、さらに、分析プロセスのどの部分までをコンピュータによって自動化するのかといった問題が生じる。これらの問題が技術の進歩によって自動的に解決されるとは考えにくい。そして、これらの問題については内容分析の分野で半世紀以上に渡って検討が行われている。少なくともこういった部分については、内容分析に学び、内容分析の分野における蓄積を継承すべきではないだろうか。

本稿では以上のような観点から、まず、内容分析の概略と定義、そして内容分析に学ぶべきと考えられる点をふり返る。その上で、内容分析の分野における蓄積を活かした日本語テキスト型データ分析方法として、新たに「計量テキスト分析」を定義し、この方法を提案する<sup>2)</sup>。これによって内容分析の分野における蓄積を継承し、発展させることを目指したい。なお本稿で主に扱うのは、上述の分析時に生じるいくつかの問題に、いかに対処すべきかという基本的な方針である。言葉を換えれば、具体的な方法というよりはむしろ、具体的な方法を背後から支える概念的基礎（フィロソフィー）の部分を中心に扱う<sup>3)</sup>。

## 2. 内容分析

### 2.1 概略と定義

#### 2.1.1 歴史と特長についての概略

内容分析と呼びうる方法が、初めて用いられた時期を正確に述べることは難しい。だが、少なくとも現在の内容分析に直接的に寄与したと思われる初期の試みとしては、19世紀から20世紀初頭のものを挙げることができる。この時期には、新聞の印刷部数が増大したことを背景に、新聞記事の計量的分析が行われるようになった。この時期に行われた分析においては、いかに「堕落した」「取るに足りない」記事に紙面が占められているのかということが主要な関心事であった（例えば Speed 1893; Mathews 1910）。その後、20世紀の前半から第二次世界大戦の前後にかけて、価値観や世論、社会的ステレオタイプといった社会学的な概念を追究するために、新聞の分析が行われるように

なった (Woodward 1934; Martin 1936)。そしてこの時期には、社会科学の理論・概念に加えて、心理学実験や市場調査のフィールドから優れた統計手法が持ち込まれ、内容分析の方法論がまとめられている (Berelson & Lazarsfeld 1948; Berelson 1952)。また第二次大戦時に、ドイツやその同盟国のマス・メディアの分析、すなわちプロパガンダ分析が大規模に行われたことも、内容分析の方法論確立に寄与したとされている (Lasswell et al. eds. 1965; George 1959)。

当初よりマス・コミュニケーション研究における利用が多いものの、いったん方法論がまとめられると、内容分析は多岐にわたる分野で用いられるようになった (Pool ed. 1959)。例えば、質問紙調査における自由回答や、心理学実験における被験者の回答の分析といった、調査の補助手段として応用されている。その他にも、心理学における小集団のコミュニケーション過程の分析や (Bales 1950)、神話や民話から共通するパターンを発見しようとする文化人類学の研究においても内容分析が利用されている (Sebeok & Orzack 1953)。また、歴史学では大量の資料を扱うために、文学では小説をはじめとする作品を扱うために内容分析が用いられたという。さらに、実務的・商業的な局面でも内容分析は利用されている。例えば CDC: Communication Development Company の試みにおいては、特定の製品やサービスについてのインタビュー調査を行い、その結果を巨大なコーパス (テキスト・データベース) に格納して、言葉の出現頻度を用いた分析を行ったという (Stone 1997)。また、McBer や Gallup といった企業においては、採用面接の評価を行うための方法として内容分析が活用されている (Spencer & Spencer 1993; Schlesinger 1993)。

このように様々な局面で内容分析が活用されていることは、以下のような内容分析の特長を考えれば、自然なことと言える。第一に、コミュニケーション活動は社会的相互作用における中心的な位置を占めているので、コミュニケーションを反映させている「内容」に注目する内容分析は、分野を問わず、多様な研究に応用することができる (Riffe et al. 1998: 31)。第二に、長期間にわたって保存・蓄積されてきた「内容」を分析対象とできることも、内容分析の特長である (Holsti 1969: 15)。例えば、価値観や社会的ステレオタイプが長期間にわたっていかに変化してきたのかといった課題に対峙するときには、この点が重要な意味を持つ。J. Z. Namewirth (1973) や W. A. Danielson & D. L. Lasorsa (1997) が行ったように 100 年前から現在に至るまでの新聞記事を収集・分析することは可能であっても、100 年前に遡って質問紙調査・繰り返し調査を実施することは言うまでもなく不可能だからだ。第三に、内容分析は調査対象者に気づかれにくい調査技術、すなわち調査される対象の反応に影響されにくい調査

技術である。言葉を換えれば、実験法や質問紙調査とは異なり、データを産み出す状況に対する過度の束縛がデータの妥当性を損なうという危険性が低い（Krippendorff 1980=1989: 35-36）。この点で、質問紙調査のような他の調査技術の補助手段としても、内容分析は有効である。例えば質問紙調査では、網羅的で完全な選択肢を提示することが難しいという選択型の設問が持つ問題（安田 1970）を、自由回答型の問い合わせを設け、得られた回答の内容分析を行うことで補うことができる。

### 2.1.2 定義

この内容分析の定義としては、例えば表1に示すようなものがある。第二次大戦の前後にいったん方法論がまとめられたとはいいうものの、その後も継続的に議論・検討が行われていることが、表1に見られる定義の変遷からうかがえるだろう。ここでは、これらの内容分析の定義とその変遷が、本稿でも受け入れられるものか、あるいは本稿でもそのまま継承できるものなのかどうかという観点から、表1を確認しておきたい。

表1: 内容分析の定義

定義	文献
The statistical semantics of political discourse.	Kaplan (1943: 230)
A research technique for the objective, systematic, and quantitative description of the manifest content of communication. (表明されたコミュニケーション内容の客観的・体系的・数量的記述のための調査技術)	Berelson (1952: 18, 1954=1957)
Any technique for making inference by objectively and systematically identifying specified characteristics of messages.	Holsti (1969: 14)
A research technique for making replicable and valid inferences from data to their context. (データをもとにそこから（それが組み込まれた）文脈に関して再現可能かつ妥当な推論を行うための1つの調査技法)	Krippendorff (1980=1989: 21)
A research methodology that utilizes a set of procedures to make valid inferences from text. These inferences are about senders, the message itself, or the audience of the message.	Weber (1985: 9)

変遷の傾向としてまず目につくことは、内容分析が単なる記述（description）の方法から、推論（inference）の方法として定義されるようになり、さらに何を推論するのかという対象についても定義に盛り込まれるようになったことだ。内容分析が扱うデータがシンボリックなもの、すなわち対象や概念をあくまで間接的に表すものであることを考えればこういった変遷は穏当なものであろう。というのも、データ中には誰もが認めるような、たった1つの「真の意味」が表明・明示されており、それを記述するのが内容分析だという考えには明らかに無理があるからだ。1つのメッセージをもとに、送り手にとっての意味や、受け手にとっての意味といった複数の意味を推論することは当然可能である。あるいは1つのデータに対して社会学的な観点や精神医学的な観点からといったように、複数の観点から解釈・推論を行うことも同様に可能である。この点を明確に示しているということに関して、表1に挙げた中でB. Berelson（1952）より後の定義は共通している。

もう1つの変遷の傾向として、統計的方法または量的方法を用いるということを明示していない定義が、近年では多くなっていることが挙げられる。確かに、量的方法を重視した定義が、質的な分析・記述を禁じているかのように解釈されることには問題がある。次節で後述するように、本稿では量的方法を用いることを前提としつつ、場合によっては質的な分析・記述をも推奨しているためだ。それでも、量的方法が分析手順の一部として含まれることを明示していないことは、本稿の立場からすれば不十分な点であり、この点は本稿における方法の説明・定義の中で明らかにしておくべき点であろう。その他には、系統的（systematic）、客観的（objective）、再現可能（replicable）、妥当（valid）といった用語、ないしはこれらの用語の組み合わせが定義によって異なることがあることが目につくかもしれない。ここでは、これらの語の定義にまで踏み込むことはせず、これらの語を用いることで、内容分析が科学的方法であることを強調している点で、表1に示した定義は概ね共通していることを指摘するにとどめたい。たった1つの「真の意味」から距離を置き、様々な観点からの解釈・推論を容認する一方で、決して好き勝手をしてよいというわけではないことを強調する意味でも、これらの用語は重要なものであろう。そして、これらの用語またはその組み合わせが必要であることは、本稿においても変わりはない。

以上に内容分析の定義をふり返ってきたが、定義の多くは穏当と考えられるものであり、本稿で継承すべき定義を1つだけ選ぶということは難しい。強いて言うならば、表1に示した定義のうち、R. P. Weber（1985）における“a set of procedures”、あるいはK. Krippendorff（1980=1989）やO. R. Holsti（1969）における“technique”と

いった部分が、本稿で提案する方法にあたる部分と言える。すなわち本稿では、"a set of procedures"ないしは"technique"と記述された部分について、新たな手順・方法を提案するのである。

## 2.2 質的データの計量的分析

### 2.2.1 計量的分析手法を用いる利点

前節で取り上げた内容分析の定義においても、比較的新しいものでは、分析が統計的・計量的なものでなければならないという記述を避けているものが多かった。また、テキスト型データのような質的データを分析する場合には、集められたデータの中から分析者が典型的だと考える事例を引用し解釈するという、質的な方法を用いることが多い。本稿でもこのような素データの引用や解釈を否定するわけではない。だが、それを行うにしても、その前の段階で計量的分析手法を用いることで、以下のような量的方法の利点を活用することが望ましいというのが本稿の立場である。

Berelson (1952) や H. D. Lasswell (1965) は、科学的でない印象批評的な方法としての「質的」分析を徹底的に批判し、量的方法の利点として、より正確で厳密な形で分析を行えることを強調した。「頻繁に」「めったに・・・ない」といった表現に頼るよりも量的方法を用いた方が、より正確かつ厳密な結果が得られる。そして、正確で厳密な数値指標を用いれば、分析結果を他の研究者による分析の結果と比較できるので、直接の比較に耐えるような研究の蓄積を築くことができる。もちろん、質的な方法による研究であっても比較は可能だが、数値指標を用いた研究の方が、より多くの研究をより正確に比較しうることは間違いない。また、例えば Lasswell (1965: 42-3) は、多くの質的分析においては、引用されている素データがたまたま研究者の目にとまったものなのか、それとも大量の資料を精査した結果まぎれもなく典型的なものとして選ばれたのかといった疑問の数々が、曖昧なままに残されていると批判している。質的データ分析における信頼性を論じた Y. S. Lincoln & E. G. Guba (1985) は、研究者がたどったデータ収集・分析の過程を他の研究者が監査 (audit) できるかどうかということを、信頼性の一部である"dependability"として取り上げている。Lasswell (1965) が指摘したような曖昧さは、この"dependability"を、ひいては信頼性を危うくするものである。以上のように、信頼性ないしは客觀性を向上させ、直接の比較検証に耐える研究を蓄積できるということが、計量的分析を行う利点の 1 つとして挙げられる。

計量的分析を行うもう 1 つの利点として、広い意味でのデータ探索が行えることがある。これは Holsti (1969: 11) が「それまでは見過ごされていたようなデータの質

的側面が、量的分析によって明らかになる場合がある（筆者訳）」と指摘している点である。言うまでもなく、統計手法は正確さ・厳密さの追求や仮説検証のためだけに利用されるものではない。J. W. Tukey (1977) の提唱した探索的データ解析 (EDA: Exploratory Data Analysis) が示すように、統計手法はデータ探索を行うための手段としても有効に利用しうる。

Holsti (1969) の述べた「質的側面」という言葉は極めて曖昧な表現であるが、いくつかの解釈が可能であろう。第一に、膨大な量のデータを扱う場合には、データ全体を記憶すること、言葉を換えれば、データを読み進めながら理解を積み重ねてデータの全体像を把握することが難しい。この状態では、データに即した分析を行うことは難しいだろう。これについては、例えば以下のような指摘がある。

自分の得たデータを概念化したり、適切なデータ集約の方法、辞書の構成、分析の手続きを決定する前に、研究者はデータの多様性、種類、分布などについての全体像を得ておきたいと思うかも知れない。

分析者が考慮すべきテクスト資料が膨大な量になる場合には、このような全体像を描くためにコンピュータを利用したいという動機づけが特に強く生じやすい。コンピュータの助けを借りなければ、分析者は偏った、不完全な、そして非常に選択的な印象を形成しがちである。

(Krippendorff 1980=1989: 188, 傍点筆者)

ここで指摘されている「データの多様性、種類、分布などについての全体像」は、明らかに、量的方法によって得られるものである。また、データの全体像や概観を描くのに適しているというのは、コンピュータそのものの利点というよりも、むしろ量的方法の持つ利点である。大量のデータに対して量的方法を適用するためにはコンピュータの助けが必要になるという観点から、Krippendorff (1980=1989) はコンピュータ利用を強調している。しかし、ここではむしろ「偏った、不完全な、そして非常に選択的な印象を形成」する前に、量的方法によってデータの全体像が得られることを強調したい<sup>4)</sup>。ここで得られるようなデータ全体を見通す視点というものを、Holsti (1969) が述べた「質的側面」の一種として捉えられるのではないだろうか。

第二に、素データから引用し解釈すべき部分が明らかになるという意味で、「質的側面」という表現を捉えることもできる。量的方法を用いた内容分析に対する批判の1つとして、少数の重要なコミュニケーションに対して質的分析を行う方が、優れた洞察を得られるという A. L. George (1959) の主張がある。また、見田宗介 (1965) はデータ中から引用・解釈すべき部分を選択する基準として「類型に関するカバレッジ」

「次元ないし要因領域に関するカバレッジ」「要因の顕現性」の3点を挙げている。しかし、多くのデータから少数の重要な部分、あるいは見田（1965）の示した基準を満たすような箇所を適切に選定することは、必ずしも容易ではない。この選定のための助けとしても、計量的手法によるデータの探索・要約が役立つ。もちろん単に要約されたデータを参考にするというだけでなく、計量的分析の結果として得られる数値を、選定に直接利用することもできる。すなわち、特異なパターンを含む部分であったり、あるいは何らかの特徴がもっとも明確に顕現している部分を、量的指標によって選択することができる。

以上に述べてきたように、計量的分析手法をテキスト型データに適用することの利点としては、大きく分けて、信頼性・客觀性の向上とデータ探索の2点を挙げることができる。伝統的な内容分析、なかでも初期の試みにおいては、このうち信頼性・客觀性に関する部分が注目され、データ探索に役立つという点は軽視される傾向があった。だが、これら2点は量的方法を用いることの利点として、いずれも同様に重要なものであろう。

### 2.2.2 量的方法と質的方法

量的分析によって、データの質的側面に関する発見が得られるという指摘からも明らかなように、内容分析においては、量的方法と質的方法とが互いに相容れない、断絶した、排他的なものであるとは捉えられていない。むしろ両者は、連続した直線上にあるもので、互いに不可分なものである（Lazarsfeld & Barton 1951）。仮に、可能な限り厳格な形で量的分析を行ったとしても、研究の様々な段階で、量的ではない質的な作業が必要になる。

まず、質的なデータを量的に扱える形に変換するという作業は、決して純粹に量的な作業ではない。通常この段階では、いくつかのカテゴリーにデータを分類していくよう、コーディングと呼ばれる作業が行われる。例えば新聞記事が、政権に対して好意的であるか非好意的であるか、あるいはアノミーに言及しているかどうかといった分類を行うのである。このコーディング作業によって、好意的な記事が多かったかどうか、アノミーに言及した記事が増減しているかどうかといった、量的分析が可能になる。このコーディング作業を実際に行う前には、いくつかのカテゴリーを作成し、特定の記述がデータ中にあればそのデータを特定のカテゴリーに分類すること、といった基準を作成する必要がある。本稿ではこの基準のことをコーディングルールと呼んでいるが、これを作成するという営みは、明らかに非量的なものである。P. F. Lazarsfeld & A. H.

Barton (1951) はこの過程を、質的カテゴリーの形成 (the formation of qualitative categories) として大きく取り上げているし、S. Saporta & T. A. Sebeok (1959) はこの過程でデータ全体を見通すような膨大な非量的作業が必要になることを指摘している。さらに、量的分析の結果がこの過程に依存していることを考えれば、D. Cartwright (1953: 424) のように内容分析とコーディングを同義語であると捉える論者がいることも無理からぬことである。

コーディングは分析における 1 つのプロセスにすぎないが、コーディングルールを作成するという作業は、研究者の社会学的想像力が發揮されるべき極めて重要な、なおかつ質的な作業である。例えば、データ中に「しわひとつない制服を着た」という表現があったときに、制服が示す権威に注目して人物の「権威主義的傾向」を表すものとして数え上げるのか、それとも単に「几帳面さ」を表すものとして数え上げるのかでは、分析の内容・結果共にまったく異なるものとなるだろう。あるいは、質問紙調査における自由回答項目で「一番大切なものは何か」をたずねた場合に、「夫」「妻」「母親」「父親」「子供」などを「家族」というカテゴリーに分類して数え上げることも考えられるし、「夫」「妻」「恋人」などを「異性愛の対象」というカテゴリーに分類して、「家族」と区別することも考えられる。このように、同じデータに向かい合う場合でも、研究者の持つ理論仮説や研究目的が異なれば、当然それにそった異なるコーディングルールが作成される。

なお、Berelson (1952) は信頼性の維持、ないしは科学的方法としての厳密性の維持のために、あくまでコミュニケーション内容の明示的な側面だけに分析対象を、ひいてはコーディングルールを限定している。だが本稿では、他の研究者がまったく同じコーディングを再現できるような形でコーディングルールを開示し、信頼性を維持するという条件のもとに、上に例示したようなコーディングルールを自由に作成して、社会学的想像力を發揮すべきだと考えている。すなわち、C. E. Osgood et al. (1957) が述べるように、コーディングルールを作成することによって、研究者の持つ理論や問題意識の操作化を目指すのである。これは質問紙調査において、質問文を作成することによって理論や問題意識の操作化が行われることと同じことである。質問紙調査においては、様々な質問文、それも時として非常に個性的な質問文によって、回答者の特定の 1 側面を拾い上げることが目指される。それと同様に、自由かつ巧妙にコーディングルールを作成することでこそ、データ中から意味ある側面を取り出し、研究目的に合致した測定を行うことができよう。以上に述べてきたような点で、量的分析を行う前の段階、すなわち質問紙調査であれば質問紙を作成する段階であり、内容分析においてはコーディング

ルールを作成する段階では、質的な作業が、それも量的分析の結果を致命的に左右するような重要な作業が必要になる。

次に、量的分析がいったん完了した後の段階を考えても、やはり量的でない質的な作業が必要になることに気づくだろう。数値をまとめた表であれ、より視覚的な形で表現されたグラフであれ、解釈を行って結論を導くのは研究者、人間である。ここでは、解釈そのものもさることながら、解釈を行うために素データを読み返したり、あるいは、行った解釈が妥当なものであるかどうかを確認するために素データにあたるといった質的な作業が行われる。また、量的方法によって、データの質的側面に関する新たな発見が得られるというのも、主としてこの段階においてのことである。

この段階で新たな発見が得られれば、量的分析を行う前の段階にまで戻って、コーディングルールの作成からやり直すということも当然起こりうる。再び質問紙を作成して調査をやり直すことが難しい質問紙調査の場合と違って、テキスト型データを分析する場合には、コーディングルールを修正して分析をやり直すという試行錯誤を行うことができる。このように質的な作業と量的な作業とが交互に行われるような試行錯誤こそ、次のような内容分析の考え方を具現化するものであろう。

質的方法が洞察にあふれ、量的方法が仮説検証のための単に機械的なものだと決めつけるべきではない。この両者の関係は循環的なものである。すなわち、それぞれが新たな洞察をもたらし、それによって他方に資するものである。

(Pool 1959: 192, 筆者訳)

### 2.3 コンピュータ利用のありかた

上に述べて来たような形で量的方法と質的方法を共に用いるためには、コンピュータを適切に利用することが望ましいだろう。単に統計計算のためというだけでなく、内容分析のためにコンピュータを用いる当初の試みとしては、「General Inquirer」(Stone et al. 1966) というプログラムがある。そして、内容分析のためにコンピュータをいかに利用すべきかという点については、P. J. Stone et al. (1966: 68-71) の指摘した点が、現在でもほとんどそのまま通用すると考えられる。

第一に Stone et al. (1966) は、コーディング作業においてコンピュータを用いる利点として、次のような指摘を行っている。コンピュータに意味の判断や文脈の理解をもとめることはできないが、「データ中に特定の表現があれば、そのデータを特定のカテゴリーに分類すること」といったコーディングルールをいったん与えてやれば、デー

タがいくら大量にあったとしても、コンピュータにコーディング作業を行わせることができる<sup>5)</sup>。また、意味や文脈を理解できないコンピュータにでも実行できるほど、明確なコーディングルールであれば、そのコーディングルールが何を測定しているのかという点に関して曖昧さが残らない。さらにコンピュータが決まり切った手順を適用してくれる所以、他の研究者の分析結果と自らの分析結果を直接比較したり、複数のコーディングルールを同じデータに適用して比較するといったことも容易に行える。

Stone et al. (1966) は信頼性または客観性という語を用いていないが、以上の点は前節で触れた信頼性・客観性に直接関わる問題である。まず、コーディングルールに曖昧さがないということは、信頼性を維持するために必須の条件である。また、統計的分析に耐えるような大量のデータを扱う場合、人手でコーディングを行うと、すべてのデータを一貫した規則のもとに分類することが容易でないという点で、コーディング作業の信頼性が問題になる。一人で大量のデータを扱えば作業が長期間にわたり、作業開始時から終了時まで一貫した規則を保つことが難しくなるし、多人数でコーディングを行っても各個人の間でコーディング規則を一致させるのは容易でないためだ。なお、労力を節約できるという点でもコンピュータはもちろん有効である。質的な作業と量的な作業とが交互に行われるような試行錯誤が可能になるのは、コーディング作業で消耗することなしに、研究者がコーディングルールの作成に集中できるからこそである。

第二に本稿では、2.2 節で前述したように量的分析の結果をもとにしてデータの質的側面を探索するというプロセスを重視している。このプロセスでは、大量のデータ中から、特定の条件を満たすデータを瞬時に検索できるというコンピュータの利点が役立つ。コンピュータを利用することで、素データの中から、量的な分析の結果が顕現しているような部分、あるいはデータの典型例や特異例を抽出するといったことが瞬時に行えるのだ。言葉を換えれば、データの質的側面を活かすための素データ検索に、コンピュータが役立つのである。この点も Stone et al. (1966) が、データへの近接性 (closeness to data) と呼んで指摘した点である。

要約すると、量的分析を行う前の段階ではコーディングの信頼性を維持し、労力を軽減することに、そして量的分析の後の段階ではデータの質的側面を活かすための検索に役立つことが、コンピュータ利用の重要な利点である。General Inquirer の作製当時とは異なり現代のコンピュータでは、コーディング処理が高速になったばかりか、検索結果をプリントアウトしなくともディスプレイ上で即座に確認できるようになった。コンピュータそのものの進歩は、上に挙げた Stone et al. (1966) の指摘を時代遅れなものにしたのではなく、むしろその有効性を高めたと言えよう。

### 3. 国内での展開

#### 3.1 国内における内容分析とコンピュータ利用

以上に量的方法と質的方法についての考え方やコンピュータ利用について、内容分析における議論を振り返りつつ、基本的にはそれらを継承するという本稿の判断・方針を述べてきた。もっとも、コンピュータを利用した分析の方法を考える際には、扱うデータが日本語であるということが大きな意味を持つ。日本語で使用される文字は15万種類以上と極めて多い上に、英文におけるスペースのような、語と語の間の明確な区切りがない。よって、英語データであれば単純な処理によって語を取り出せるので、頻出語リストを自動作成することも容易だが、日本語データに対して同様の処理を行うためには、形態素解析のような、自然言語処理の分野における近年の成果を待たねばならなかつたのである（長尾編1996）。こういった言語の壁を考えれば、これまでの日本における研究をふり返ることも重要である。

本稿の冒頭で触れたように、内容分析の国内での普及・発展は必ずしも十分とは言い難く、コンピュータを用いた内容分析の試みとなるとますますその数は限られたものとなる。外務省大臣官房電子計算機室（1967）においては、2.3節でも触れた "General Inquirer" が紹介されており、独自のプログラム開発を行う予定である旨の記述もあるが、それ以後の成果は公表されていない。また新聞記事データベースが整備されたことをきっかけに、それを内容分析に利用する方法についての検討が行われたが、これは主としてデータ収集法に関する検討であり、コーディングや分析にまでコンピュータを利用するというものではない（河野1992, 1993; 三上1994; 竹下1994）。その他には宮崎正也（2001）において、（株）構造計画研究所の Content Analyzer という Macintosh 用分析プログラムを内容分析に用いたとの記述がある。このプログラムは、文章から自動的に言葉を取り出してカウントしたり、取り出した言葉を分類するという形でコーディングを行えるプログラムのようであるが、現在では販売されておらず、これ以上の詳細はわからない。こうした中で、吉田文彦による一連の研究は目を引くものがある（吉田1997, 1998, 2000, 2002）。しかし、残念ながらこれら一連の研究は「内容分析のため」とされているものの、現時点では形態素解析や構文解析といった技術開発の色彩が濃いようであり、これらの技術を用いた実際の内容分析のあり方については、いまだ十分に明確とは言い難い。以上に挙げた研究はいずれも総じて先駆的なものではあるが、本稿で直接的に継承・発展できるような性質のものではない。

むしろ、必ずしも内容分析という語が使われていないような分野で、比較的本稿の方針に近い形の先行研究を見ることができる。例えば教育学ないしは教育工学という分野では、学校で行われた授業の逐語記録を分析する方法について、検討が行われている。なかでも興味深いのは、コンピュータを用いて授業の逐語記録から機械的に言葉を取り出した上で<sup>6</sup>、発言回数・発言量・エントロピーなどの指標を算出したり、頻出する言葉が授業全体の中のどの部分で出現しているのかを確認するといった方法が試みられていることだ（大谷・松原 1984; 大谷 1989; 前迫ほか 1990）。また、機械的に取り出した言葉を多次元尺度構成法によって布置したり、クラスター分析によって分類したりといった形で、多変量解析も応用されている（柴田 1995）。ここでは「量的な分析によって質的分析に取って代わろうとするものではなく、大谷と同様にむしろ全体としては質的な授業分析を支援するものとして、量的な手法をとりいれ」（柴田 1996: 217）ことが目指されている。これらの研究は、新たな方法を様々なデータに対して適用し、十分に方法を確立するという段階にまではいたっていないようだが、量的方法と質的方法を相補的に用いるという本稿の方針（2.2 節）に合致する先行研究と言えよう。

次に社会学における社会調査法の分野でも、現在では内容分析という語があまり使われていないものの、原純輔（1988）や原編（1992）をはじめとして非定型データの分析方法に関する検討が活発に行われている<sup>7</sup>。なかでも、質問紙調査における自由回答項目の分析法が検討される中で、2.3 節でコンピュータの重要な用途の 1 つとして触れたような形、すなわち、コーディング作業を自動化するという形で、コンピュータが利用されている。ここでは、都築一治（1992）や高橋和子（2000）のように職業をたずねた自由回答項目への記述を分類するという、特定のコーディングルールに特化した研究と、佐藤裕（1992）や田中重人・太郎丸博（1996）のように汎用的なコーディングルールを扱うための研究がある。自由回答のコーディングにコンピュータを用いるというこれらの研究は、その後、川端亮を中心に発展的に継承され、分析の中の 1 プロセスであるコーディングにとどまらず、分析プロセス全体をドライブする方法論として「計量テキスト分析」が提案されている（川端 1997, 2001, 2003; 川端編 1999, 2004; 秋庭・川端 2004）。この計量テキスト分析について、節をあらためて述べる。

### 3.2 計量テキスト分析

この研究は、もともと質問紙調査における自由回答項目の分析方法としてスタートしたが、既に川端編（1999）がまとめられた時点で、自由回答だけにとどまらず新聞記事や雑誌記事、宗教教典等が分析対象となっている。さらに川端（2001）ではインタ

ビューデータの分析も加わり、現在ではテキスト型データ一般を扱う方法となっている。またこの方法は、もともとはコンピュータを用いてコーディングを行うという部分に重点が置かれており、「コンピュータ・コーディング」という名称で呼ばれていたが（川端 2001）、次第に、量的方法による探索という側面が重視されるようになる（川端 2003）。そして、量的方法の利点を活かした方法であることを強調し、単なる「コンピュータ・コーディング」という言葉と区別するために、計量テキスト分析という名称が考案されている。そして、この方法がもっとも簡潔に述べられている記述としては、以下のようなものがある。

計量テキスト分析とは、インタビューデータなどの質的データ（文字データ）をコーディングによって数値化し、計量的分析手法を適用して、データを整理、分析、理解する方法である。

（秋庭・川端 2004: 235-6）

なお、同じように計量テキスト分析と呼ばれている方法であっても、質問紙調査における自由回答を扱う場合と、インタビューデータを扱う場合とでは、具体的な手順・方法にはいくぶん違いがある。よって、計量テキスト分析という名称は、1つの具体的な分析手順・方法を指すものというよりも、量的方法の利点を活かしたテキスト型データ分析方法の総称であると読み取ることができる。

この計量テキスト分析という方法には、内容分析に依拠したものである本稿の方針と、多くの点で共通する部分がある。例えば、単に信頼性を向上させるというだけでなく、以下のようにデータ探索を行うという目的で量的方法を用いている点がそうだ。

計量分析することによって、単なる自由回答やテキストデータを読んでいるだけでは気づかない、あるいは気づきにくいデータの「潜在的論理」を発見できる可能性があると考える。分析の対象とする文章を読んで、文字通りに解釈するのではなく、文をいったん語に分解し、その語と語の間の連関の強さを元に単語の使われる暗黙の意味構造を新たな「潜在的論理」として、いわば文法のように取り出し、データ理解の枠組みとして用いるのである。つまり計量的な分析が、多変量解析の手法を用いて潜在的な概念を見いだすのと同じようなことが、質的データを対象として実現可能となるはずである。これが実現すれば、文章を読んでいるだけではわからない分析のための新たな発想、発見を得ることができることがある。

（川端 2003: 41）

ここでは同時に、言葉を細かな文脈からいいたん切り離して集計することで、インタビューを実施するときに要求されるような共感的・共同的な視点から距離を置いた、より冷静にデータを眺める視点を得られるという利点も指摘されている<sup>8)</sup>。なお、量的方法によって、テキスト型データに潜在する論理を取り出すという考え方は、川端編（1999）においても記述されており、計量テキスト分析において一貫して追求されてきた事柄である。そして、ごく最近では内容分析の分野においても、量的方法で文章から潜在的な意味を取り出すという似通った発想が見受けられる。というのも、近年の認知科学の分野では、自然言語を理解するために、人の脳で多変量解析（因子分析）と似通ったプロセスが用いられているとして、そのプロセスを模倣して言葉の意味を探る方法 "Latent Semantic Analysis (LSA)" が提案されている (Landauer & Dumais 1997)。そしてごく最近になって A. F. Simon & M. Xenos (2004) が、この LSA を内容分析に応用することを試みている<sup>9)</sup>。

さらに川端（2003）は分析やコーディングにおける試行錯誤を、以下のように重要なものと見なしている。

そして現代の高性能のコンピュータによって、これらの道具を使った分析が、繰り返し試行錯誤することが可能になった。単語の統合の仕方を何度も変え、そのたびごとにクロス表やコーディングルールファイルを作り直し、等高線図を書き直すことが、短時間でできることによってこそ、試行錯誤は可能となる。きちんとした手順の中で、反復し、試行錯誤し、確かめるところに新たな発見、思いつきが生まれる。

(川端 2003: 49)

このような試行錯誤はまさに、2.2 節で触れた、質的な作業と量的な作業とが交互に行われるような試行錯誤に他ならない。さらに計量テキスト分析においては、量的方法と質的方法を共に用いて互いを補うような試行錯誤の結果、質的な分析・記述に重点を置いて報告がなされる場合がある。これは特に、インタビューデータを扱う場合に多く行われることだ。量的な分析をいったん行うことで、量的な方法ではすくいとることが難しい部分、すなわち質的データのおもしろさと呼ばれるような部分がより明確になる。よって、質的データのおもしろさを活かすためにも、量的分析の結果を利用できるということが指摘されている (川端 2001)。そして、量的分析の結果を参考として質的な記述を行うという手順・方法も、計量テキスト分析には用意されているのである (川端 2003: 168-70)。

この点は、内容分析に対する、計量テキスト分析の大きな特長である。というのも 2.

2 節で触れた、量的方法と質的方法を共に用いて互いを補うという I. d. S. Pool (1959) の考え方は、内容分析において十分に受け継がれてきたとは言い難い。内容分析の概説書を見ても、場合によっては、ほとんど無視されているようにさえ見える。そうでない場合にしても、例えば Holsti (1969) は明示的にこの考え方を採用したが、実際に分析を行う方法として彼が提示したものは、量的方法に大きく偏ったものだった。分析の結果として報告・提示されるもっとも重要な部分は、もっぱら量的な分析の結果だったのである。量的方法と質的方法を共に用いて互いを補うという考え方、あるいは、量的方法と質的方法が断絶したものではなく連続したものだという考え方を、計量テキスト分析では、内容分析において以上に重視しており、また分析の実践に活かそうとしているのである。この点で計量テキスト分析は、単に日本語を分析対象とした先行研究であるにとどまらず、日本語・英語という言語の問題を越えた方法論的示唆を持つ研究である。

#### 4. 計量テキスト分析の拡充

##### 4.1 定義

計量テキスト分析は前節に述べたような点で優れた方法だが、これまでのところ、内容分析の分野における蓄積が十分には取り入れられていない点に、精緻化ないし拡充の余地が残されていた。分析を行う際の基本的な考え方ないし概念的基礎（フィロソフィー）の部分だけでなく、実際に分析を行うための具体的な手順の部分においても、内容分析の考え方を取り入れることは有益だと考える。こういった改善によって、内容分析の分野における蓄積を実際の分析の中で活用できるようになれば、単に計量テキスト分析の拡充というだけでなく、内容分析の発展にも寄与しうるのではないだろうか。こうした観点から精緻化・拡充を行った計量テキスト分析の方法について、その概観を以下に示し、この方法を提案することで本稿を締めくくりたい。

まず、分析がいかに行われるべきかという基本的な考え方をできるだけ簡潔に表すために、本稿では計量テキスト分析を次のように定義する<sup>10)</sup>。

計量テキスト分析とは、計量的分析手法を用いてテキスト型データを整理または分析し、内容分析（content analysis）を行う方法である。  
計量テキスト分析の実践においては、コンピュータの適切な利用が望ましい。

第一に、この定義では「内容分析（content analysis）を行う方法」という言葉によって、計量テキスト分析を内容分析の一種または一部として位置づけている。すなわち

2.1 節の末尾で触れたように、内容分析の定義における "a set of procedures" ないしは "technique" という部分の具体的手順・方法を提供するものとして計量テキスト分析を位置づけている。これによって、分析方法が信頼性・妥当性を備えねばならないことや、単なるデータの記述ではなく推論を含むといった、内容分析の概念的基礎を受け継ぐことを明示している<sup>11)</sup>。

第二に、この定義では「計量的分析手法を用いてテキスト型データを整理または分析」という言葉によって、量的方法を用いることを明示している。もちろん2.2 節で触れたように、ここで言う量的方法とは、必然的に質的方法を含むものである。さらに前節で触れたように、量的方法と質的方法とを共に用いて互いを補った結果として、量的分析の結果はデータの「整理」にとどまり、質的な分析・記述が主に報告される場合もあるだろう。また量的分析の結果が、主たる「分析」の結果として報告される場合ももちろんあるだろう。こういった意味で「整理または分析」という言葉を用いた。

第三に、2.3 節で触れたように、計量テキスト分析を行うにあたってはコンピュータを利用するすることが望ましい。定義においてコンピュータ利用を必須の条件とするのを避けたのは、コンピュータを利用せずに計量テキスト分析を行うことが絶対に不可能であるとまでは言い切れないためだ。コンピュータ無しに計量テキスト分析を行うことは決して容易ではないだろうが、計量テキスト分析の方法としての特長は、コンピュータ利用というよりも、上述の量的方法と質的方法の組み合わせという点にある。

#### 4.2 具体的な方法

定義の中で「整理または分析」という言葉を使ったが、このうち「整理」の側によった分析方法とは、例えばインタビューデータを対象として量的な方法でデータを探索し、その結果を参考にして質的な記述を主に行うという方法である。この方法には、素データの引用や解釈、あるいは質的な記述を行うという手順が含まれており、そういう部分では内容分析というよりもグラウンデッド・セオリー・アプローチ (Glaser & Strauss 1967=1996; Glaser 1978; Strauss 1987; Strauss & Corbin 1990=1999) の考え方方がより直接的に取り入れられている。この点は、グラウンデッド・セオリーの影響を受けた部分と、計量テキスト分析がグラウンデッド・セオリーとは異なっている部分とを整理するという形で、川端 (2003) に詳述されている。この方法の具体的な手順については、川端 (2003) の他に川端 (2001) が詳しい。さらに、その方法を実現するためのソフトウェアとしても、川端と谷口敏夫によって制作された分析用プログラム「KT2システム」(谷口 1999, 2004) が用意されている<sup>12)</sup>。

次に「分析」の側によった方法、すなわち、量的な分析の結果が重視される場合の具体的な分析方法については、樋口耕一（2004, 2005）にまとめられており、方法を実現するためのソフトウェアとしても「KH Coder」を作成・公開している<sup>13)</sup>。この分析方法では、内容分析の考え方をできる限り取り入れることが目指されている。というのも、量的な分析の結果を重視する方法となると、内容分析の分野に非常に多くの蓄積があるためだ。その中には、自動的にデータ中から言葉を取り出し、多変量解析による探索を行うといった1960年代に試みられた方法、現在ならばテキストマイニングと呼ばれることが多いような方法も含まれている。「分析」よりの方法とソフトウェアは、内容分析の具体的な方法をより積極的に取り入れ、より大規模なデータの計量的分析を想定している点で、「整理」よりのものと異なっている。

なお、ここで注意が必要なのは、「整理」と「分析」が決して排他的なもの、あるいは分離したものではないことである。例えば、「分析」の側によった方法を用いる場合でも、素データを例示しようとする際には、「整理」の方法を利用することができる。すなわち川端（2003: 168-70）が提示した、量的分析の結果を参考にしながら質的な記述を行うための手順・方法を利用することができる。また逆に、「整理」の側によった方法でインタビューデータを分析するにしても、データの量が多い場合には、量的方法を用いる際に「分析」よりの方法、ないしはKH Coderを有効に利用できる。このように、「整理」を行うにせよ「分析」を行うにせよ、互いの手順・方法を部分的に共有することができる原因是、分析方法の基盤をなすフィロソフィーが共通しているからである。共通している点の中でも特に重要なのは、「整理」においても「分析」においても、量的方法と質的方法を共に用いて互いを補おうとしていることだろう。「整理」と「分析」は、別々のまったく異なる方法ではなく、共通する概念的基礎（フィロソフィー）のもとに、互いの手順・方法を部分的に共有する、計量テキスト分析のバリエーションである<sup>14)</sup>。

本稿では、内容分析を行うための方法として計量テキスト分析を定義した。2.1節で触れたように、内容分析は極めて多様な形で応用されてきた調査技法であるから、計量テキスト分析もそれに準じて、応用範囲の広い方法であると考えられる。さらに計量テキスト分析では、伝統的な内容分析よりも柔軟に質的方法と量的方法とを組み合わせられるようになったため、応用の可能性が一層広がったのではないだろうか。また、分析を行うための理論仮説や問題意識、分析時の試行錯誤といったものを完全に省くことはできないが、分析用のソフトウェアが整備されてきたことで、単純作業による消耗を避けられるようになった。さらに、やみくもな試行錯誤ではなく、きちんとした手順の中

で試行錯誤を行う方法であるということも、実用性を高めていよう。少なくとも「手数がかかるわりに報いられる点が少ない」という印象を否定できる程度には実用的で応用範囲の広い方法として、ここに計量テキスト分析を提案したい。

## 付記

本稿は樋口（2005）の第1章を中心に、編集を加えたものである。また、本稿は科学研究費補助金による研究成果の一部である（特別研究員奨励費「計量テキスト分析による情報行動の調査研究」研究代表者：樋口耕一）。

## 注

- 1) 標準番号（ISBN コード）を付され、書店に流通している書籍に関して。
- 2) 質的データの中でも特にテキスト型データを扱うのは、当初に取り上げられた新聞をはじめとしてこれまでに多くの分析が行われていることと、技術的な問題として、2.3 節で後述するようなコンピュータ処理が容易だことがある。これまでの分析例が多くあるテキスト型データを対象として方法を確立しつつ、技術的な問題の解決をみることができれば、いずれは音声・映像にも応用範囲を広げていきたい。
- 3) 計量テキスト分析の具体的な方法と、その方法を実現するためのソフトウェアについては4.2 節に挙げる文献や Web サイト等をご参照されたい。
- 4) もちろん2.3 節で後述するように、コンピュータ利用の利点の1つとして、量的方法を用いるための助けになるという点は本稿でも重視している。だが、コンピュータは極めて多様な使い方が考えられる道具であることから、本稿では何のために、またどのようにコンピュータを利用することが有効であるのかを極力明確に論じることを目指している。そこで、この節では量的方法の利点についてまとめ、2.3 節で、量的方法を用いる助けとして、コンピュータをどのように役立てられるのかを述べることとした。
- 5) Stone et al. (1966)においては「コーディングルール」ではなく、「category systems」という用語が用いられている。本稿では用語の混乱を極力避けるために、できるだけ「コーディングルール」で統一することとした。
- 6) ただし、ここで挙げた研究の中で大谷尚・松原伸一（1984）については言葉を取り出す際にコンピュータを用いておらず、取り出しの方針を定めた上で、実際の作業は国語科の教師に依頼するという方法が取られている。
- 7) 質問紙調査の結果として得られるデータは通常、n 行 m 列の行列形式となり、行

が個々のケース（回答者）にあたり、列が変数（質問）にあたるものとなる。このような一般的な形を取らないデータが、「非定型データ」と呼ばれている。

- 8) ここで川端（2003）は、グラウンデッド・セオリー・アプローチにおいて「データの切片化」と呼ばれている事柄を、自らの方法に援用している。
- 9) 2.2 節で述べたように Holsti （1969）は、量的分析によってデータの「質的側面」に関する発見が得られることがあると指摘している。ここで触れた、テキスト型データに含まれる「潜在的論理」あるいは潜在的な意味や構造といったものは、この「質的側面」の一種として捉えることができる。
- 10) これまでのところ、川端はこの方法に名称を与えたものの、定義は行っていない。
- 11) 少なくとも筆者の解釈では、このような内容分析の概念的基礎を受け継ぐことは、川端（2003）の述べる計量テキスト分析と一切矛盾しない。
- 12) KT2 システムは無償公開されており、現在は次の URL から入手することができる。  
<http://www.koka.ac.jp/taniguti96M/0/20/KTCoder/KTCoder.html>
- 13) KH Coder はソースコードの閲覧・編集が可能なフリー・ソフトウェアとして無償公開されており、現在は次の URL から入手することができる。<http://khc.sourceforge.net>
- 14) 「整理」の側によった方法を用いるか、それとも「分析」の側によった方法を用いるのかは、主として研究目的や、データの性質・量によって決定されることだろう。質的な分析・記述が必要となる研究目的のもとでは、当然、量的方法に過度に重点を置いたとしても得られるものは少ない。また、データの量が少ない場合や、統計的検定が効果を發揮しにくいような形で収集されたデータについても同様である。そういういったことに加えて、分析によって得られた発見・結論がいかなるものであるか、すなわち、それを記述するのに質的方法が適しているのか量的方法が適しているのかといったことによって、最終的に分析が「整理」と「分析」のどちらの側にどの程度傾斜するかが決まるだろう。

## 文献

- 秋庭裕・川端亮, 2004, 『靈能のリアリティへ－社会学、真如苑に入る』新曜社.
- Bales, R. F., 1950, *Interaction Process Analysis*, Reading: Addison-Wesley.
- Berelson, B., 1952, *Content Analysis in Communication Research*, New York: Free Press.
- Berelson, B. & P. F. Lazarsfeld, 1948, *The Analysis of Communication Content*, Chicago: Chicago University Press.
- Cartwright, D., 1953, "Analysis of Qualitative Material," L. Festinger & D. Katz eds., *Research*

- Methods in the Behavioral Sciences*, Niles, IL: Dryden, 421-70.
- Danielson, W. A. & D. L. Lasorsa, 1997, "Perceptions of Social Change: 100 Years of Front-Page Content in the New York Times and the Los Angeles Times," C. W. Roberts ed., *Text Analysis for the Social Sciences*, Mahwah: Lawrence Erlbaum, 103-15.
- Franke, J., G. Nakhaeizadeh, & I. Renz eds., 2003, *Text Mining: Theoretical Aspects and Applications*, Heidelberg: Physica-Verlag.
- 外務省大臣官房電子計算機室, 1967, 『CONTENT ANALYSIS —電子計算機による文章容分析』外務省.
- George, A. L., 1959, *Propaganda Analysis: A Study of Inferences Made from Nazi Propaganda in World War II*, Row: Peterson.
- Glaser, B. G., 1978, The Theoretical Sensitivity: Advances in the Methodology of Grounded Theory, Mill Valley: The Sociology Press.
- Glaser, B. G. & A. L. Strauss, 1967, *The Discovery of Grounded Theory: Strategies for Qualitative Research*, Chicago: Aldine. (=1996, 後藤隆・大出春江・水野節夫訳『データ対話型理論の発見 —調査からいかに理論をうみだすか』新曜社。)
- 原純輔, 1988, 「非定型データの処理・分析」海野道郎・原純輔・和田修一編『数理社会学の展開』数理社会学研究会, 461-72.
- 原純輔編著, 1992, 『非定型データの処理・分析法に関する基礎的研究』平成3年度文部省科学研究費補助金（総合研究A）研究成果報告書.
- 樋口耕一, 2004, 「テキスト型データの計量的分析－2つのアプローチの峻別と統合」『理論と方法』19(1) : 101-15.
- , 2005, 「計量テキスト分析の方法と実践」大阪大学大学院 人間科学研究科 平成16年度博士論文.
- 日吉昭彦, 2004, 「内容分析研究の展開」『マス・コミュニケーション研究』64: 5-24.
- Holsti, O. R., 1969, *Content Analysis for the Social Science and Humanities*, Reading: Addison-Wesley.
- 藤井美和・小杉考司・李政元, 2005, 『福祉・心理・看護のテキストマイニング入門』中央法規出版.
- 池内一, 1954, 「内容分析の方法について（上）」『東京大学新聞研究所紀要』3: 33-46.
- , 1956, 「内容分析の方法について（下）」『東京大学新聞研究所紀要』5: 47-58.
- 石井哲, 2002, 『テキストマイニング活用法－顧客志向経営を実現する』リックテレコム.
- 加納恒男, 1961, 「安保反対闘争記事の内容分析」『新聞研究』119.
- Kaplan, A., 1943, "Content analysis and the theory of signs," *Philosophy of Science*, 10: 230-47.
- 川端亮, 1997, 「パソコンによるテキストデータの内容分析」『ソシオロジ』42(1) : 97-103.
- , 2001, 「コンピュータ・コーディングによる宗教的ライフヒストリーの記述」『宗教と社会』7: 133-54.
- , 2003, 「宗教の計量的分析—真如園を事例として」大阪大学大学院 人間科学研究科 平成14年度博士論文.
- , 2004, 「計量的テキスト分析」『社会調査における非定型データ分析支援システムの開発』1-12.

- 川端亮編著, 1999, 『非定型データのコーディング・システムとその利用』平成8年度~10年度科学研究費補助金(基盤研究(A)(1)) (課題番号 08551003) 研究成果報告書, 大阪大学。
- , 2004, 『社会調査における非定型データ分析支援システムの開発』平成13年度~15年度科学研究費補助金(基盤研究(B)(2)) (課題番号 13410049) 研究成果報告書, 大阪大学。
- 河野武司, 1992, 「新聞記事データベースによるPKO報道の内容分析」『民主主義研究会紀要』21: 91-115.
- , 1993, 「国内新聞記事データベース・サービスと内容分析－PKO報道を事例として」『杏林社会科学研究』9(2): 37-58.
- Krippendorff, K., 1980, *Content Analysis: an Introduction to its Methodology*, London: Sage.  
(=1989, 三上俊治ほか訳『メッセージ分析の技法－「内容分析」への招待』勁草書房.)
- Landauer, T. K. & S. Dumais, 1997, "A Solution to Plato's Problem: The Latent Semantic Analysis Theory of the Acquisition, Induction, and Representation of Knowledge," *Psychological Review*, 104: 211-40.
- Lasswell, H. D., 1965, "Why be quantitative?", H. D. Lasswell et al. eds., *Language of Politics*, Cambridge: MIT Press, 40-52.
- Lasswell, H. D. et al. eds., 1965, *Language of Politics*, Cambridge: MIT Press.
- Lazarsfeld, P. F. & A. H. Barton, 1951, "Qualitative Measurement in the Social Sciences, Classification, Typologies, and Indices," D. Lerner & H. D. Lasswell eds., *The Policy Sciences: Recent Developments in Scope and Method*, Stanford: Stanford University Press, 180-8.
- Lincoln, Y. S. & E. G. Guba, 1985, *Naturalistic Inquiry*, Beverly Hills, CA: Sage.
- 前迫孝憲・多胡賢太郎・清水康敬・坂本昂, 1990, 「授業分析におけるコンピュータの利用の検討」『日本教育工学雑誌』14(2): 81-8.
- 真鍋一史, 1974, 「マス・コミュニケーションの調査－新聞記事の内容分析」『関西学院大学 社会学部紀要』28: 15-36.
- Martin, H., 1936, "Nationalism and Children's Literature," *Library Quarterly*, 6: 405-18.
- Mathews, B. C., 1910, "A Study of a New York Daily," *Independent*, 68: 82-6.
- 三上俊治, 1988, 「放送メディアの内容分析－その方法論的考察」『放送学研究』38: 101-18.
- , 1994, 「新聞記事データベースの有効活用法に関する研究」『パーソナル文献データベース構築の手法と評価に関する研究』東洋大学社会学研究所研究報告集(16): 58-98.
- 見田宗介, 1965, 「『質的』データ分析の方法論的諸問題」『社会学評論』15(4): 79-41.
- 宮崎正也, 2001, 「内容分析の企業行動研究への応用」『組織科学』35(2): 114-27.
- 長尾真編, 1996, 『自然言語処理』岩波書店.
- Namenwirth, J. Z., 1973, "Wheels of Time and the Interdependence of Value Change in America," *Journal of Interdisciplinary History*, 3: 649-83.
- 大谷尚, 1989, 「授業研究の一手法としての逐語記録に対する計量的な分析について」『名古屋大学教育学部紀要』36: 327-38.
- 大谷尚・松原伸一, 1984, 「出現語の頻度分布にもとづく授業の特徴化について」『電子通信

- 学会技術研究報告』84(8) : 1-6.
- Osgood, C. E., G. J. Suci, & P. H. Tennenbaum, 1957, *The measurement of Meaning*, Urbana: University of Illinois Press.
- Pool, I. d. S., 1959, "Trend in Content Analysis Today: A summary," *Trends in Content Analysis* 189-234.
- Pool, I. d. S. ed., 1959, *Trends in Content Analysis*, Urbana: University of Illinois Press.
- Riffe, D., S. Lacy, & F. Fico, 1998, *Analyzing Media Messages: Using Quantitative Content Analysis in Research*, London: Lawrence Erlbaum Associates.
- Roberts, C. W. ed., 1997, *Text Analysis for the Social Sciences*, Mahwah: Lawrence Erlbaum.
- Saporta, S. & T. A. Sebeok, 1959, "Linguistic and Content Analysis," I. d. S. Pool ed., *Trends in Content Analysis*, Urbana: University of Illinois Press, 131-50.
- 佐藤裕, 1992, 「自由回答のコンピュータコーディング」『第14回数理社会学会大会研究報告要旨集』.
- Schlesinger, L., 1993, "How to Hire by Wire," *Fast Company*, 1: 86-91.
- Sebeok, T. A. & L. H. Orzack, 1953, "The Structure and Content of Cheremis Charms," *Anthropos*, 48: 369-88.
- 柴田好章, 1995, 「概念の出現パターンに着目した授業の量的分析研究—小学校6年社会科『沖縄戦』の授業の分析」『名古屋大学教育学部紀要』42(1) : 185-95.
- , 1996, 「逐語記録にもとづく授業分析の諸手法の検討 ——質的分析への量的分析の統合をめざして」『名古屋大学教育学部紀要』43(2) : 217-28.
- Simon, A. F. & M. Xenos, 2004, "Dimensional Reduction of Word-Frequency Data as a Substitute for Intersubjective Content Analysis," *Political Analysis*, 12(1) : 63-75.
- Speed, G., 1893, "Do Newspapers now give the news?," *Forum*, 15: 705-11.
- Spencer, L. M. & S. M. Spencer, 1993, *Competence at Work: Models for Superior Performance*, New York: Wiley.
- Stone, P. J., 1997, "Thematic Text Analysis: New Agendas for Analyzing Text Content," C. W. Roberts ed., *Text Analysis for the Social Sciences*, Mahwah: Lawrence Erlbaum, 35-54.
- Stone, P. J., D. C. Dunphy, & D. M. Ogilvie, 1966, *The General Inquirer: A Computer Approach to Content Analysis*, Cambridge: MIT Press.
- Strauss, A. L., 1987, *Qualitative Analysis for Social Scientist*, Cambridge: Cambridge University Press.
- Strauss, A. L. & J. Corbin, 1990, *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*, London: Sage. (=1999, 南裕子監訳『質的研究の基礎 —— グラウンデッド・セオリーの技法と手順』医学書院.)
- 高橋和子, 2000, 「自由回答のコーディング支援 —— 格フレームによるSSM職業コーディング自動化システム」『理論と方法』15(1) : 149-65.
- 竹下俊郎, 1994, 「内容分析のツールとしての新聞記事データベース」『新聞研究』516: 60-3.
- 田中重人・太郎丸博, 1996, 「文章データのコンピュータコーディング —プログラムとその応用」『第22回数理社会学会大会研究報告要旨集』, 20-3.
- 谷口敏夫, 1999, 「全文からの『位置情報付き用語』の抽出」川端亮編著『非定型データのコーディング・システムとその利用』, 平成8年度~10年度科学研究費補助金(基盤研

- 究（A）（1）（課題番号 08551003）研究成果報告書,大阪大学, 31-58.
- , 2004, 「KT2の利用法：犯人の可視化 —小栗虫太郎『聖アレキセイ寺院の惨劇』より」川端亮編著 『社会調査における非定型データ分析支援システムの開発』, 平成13年度～15年度科学研究費補助金（基盤研究（B）（2））（課題番号 13410049）研究成果報告書, 大阪大学, 13-34.
- 都築一治, 1992, 「職業コーディング自動化システムの試験的構築」原純輔編著 『非定型データの処理・分析法に関する基礎的研究』, 平成3年度文部省科学研究費補助金（総研究A）研究成果報告書, 205-14.
- Tukey, J. W., 1977, *Exploratory Data Analysis*, Reading: Adison-Wesley.
- Weber, R. P., 1985, *Basic Content Analysis*, Beverly Hills: Sage.
- Woodward, J. L., 1934, "Quantitative Newspaper Analysis as a Technique of Opinion Research," *Social Forces*, 12: 526-37.
- 山崎秀夫・松田潤, 2002, 『顧客を創造するテキストマイニング』日本工業新聞社.
- 安田三郎, 1970, 『社会調査の計画と解析』東京大学出版会.
- 吉田文彦, 1997, 「内容分析のための日本語形態素解析システムの構築」『東海大学紀要 文学部』68: 46-56.
- , 1998, 「内容分析のためのニューラル・ネットワーク・モデルによる日本語構文解析システム構築の試み」『東海大学紀要 文学部』70: 64-74.
- , 2000, 「日本語の報道記事を対象とする構文解析システムの精度改善の試みと事象データ抽出システムの試作結果」『東海大学紀要 文学部』74: 59-72.
- , 2002, 「日本語の報道記事を対象とする事象データ抽出システム —コンピューによる内容分析手法の概要と解析精度」『東海大学紀要 文学部』77: 114-32.
- , 2004, 「内容分析研究におけるコンピュータの利用状況 —米国のこと例を中心として」『マス・コミュニケーション研究』64: 41-69.

## From Content Analysis to Quantitative Text Analysis

Koichi HIGUCHI

Content Analysis (CA) is a method used in the analysis of qualitative data acquired through social research. Though in Japan this method was first introduced and utilized in 1950s, it did not gain great popularity thereafter. In contrast, an automated and computerized method called Text Mining, which was first used in the field of information science, is quite popular now.

However, it would be a critical mistake to throw out the fruits of CA studies. Unlike the automatic summarization process, research decisions and human intelligence are essential to the data analysis process. We cannot avoid such issues as (i) how to maintain reliability or objectivity, (ii) how to balance the advantages and disadvantages of quantitatively analyzing qualitative data, and (iii) what part of the data analysis process should be automated and to what degree. It is difficult to imagine that advancements in computer technologies will automatically resolve these matters, matters that have been discussed for over half a century in the field of CA. Therefore, it would be useful to learn from such discussions.

In this light, the author first reviews the history and characteristics of CA, and points out what we can learn from methodological studies regarding CA. Then, the author defines and proposes a new method Quantitative Text Analysis which exploits both the basic ideas of CA and today's advanced computer technologies.